UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
10/684,773	10/14/2003	Zicheng Liu	M61.12-0547	3356

27366 7590 02/23/2007
WESTMAN CHAMPLIN (MICROSOFT CORPORATION)
SUITE 1400
900 SECOND AVENUE SOUTH
MINNEAPOLIS, MN 55402-3319

EXAMINER	
WOODS, ERIC V	

ART UNIT	PAPER NUMBER
2628	

SHORTENED STATUTORY PERIOD OF RESPONSE	MAIL DATE	DELIVERY MODE
3 MONTHS	02/23/2007	PAPER

Please find below and/or attached an Office communication concerning this application or proceeding.

If NO period for reply is specified above, the maximum statutory period will apply and will expire 6 MONTHS from the mailing date of this communication.

Office Action Summary	Application No.	Applicant(s)
	10/684,773	LIU ET AL.
	Examiner	Art Unit
	Eric Woods	2628

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) OR THIRTY (30) DAYS, WHICHEVER IS LONGER, FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

1)☒ Responsive to communication(s) filed on *08 December 2006*.

2a)☐ This action is **FINAL**. 2b)☒ This action is non-final.

3)☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

4)☒ Claim(s) *1,2,6,7,10,13-29,33-39 and 43-47* is/are pending in the application.

4a) Of the above claim(s) _____ is/are withdrawn from consideration.

5)☐ Claim(s) _____ is/are allowed.

6)☒ Claim(s) *1,2,6,7,10,13-29,33-39 and 43-47* is/are rejected.

7)☐ Claim(s) _____ is/are objected to.

8)☐ Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

9)☐ The specification is objected to by the Examiner.

10)☐ The drawing(s) filed on _____ is/are: a)☐ accepted or b)☐ objected to by the Examiner.

Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).

Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).

11)☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

Priority under 35 U.S.C. § 119

12)☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

a)☐ All b)☐ Some * c)☐ None of:

1.☐ Certified copies of the priority documents have been received.

2.☐ Certified copies of the priority documents have been received in Application No. _____.

3.☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

* See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

1) ☒ Notice of References Cited (PTO-892)
2) ☐ Notice of Draftsperson's Patent Drawing Review (PTO-948)
3) ☐ Information Disclosure Statement(s) (PTO/SB/08) Paper No(s)/Mail Date _____.
4) ☐ Interview Summary (PTO-413) Paper No(s)/Mail Date. _____.
5) ☐ Notice of Informal Patent Application
6) ☐ Other: _____.

DETAILED ACTION

Response to Arguments

Applicant's arguments, see Remarks pages 1-5 and claim amendments, filed 12/08/2006, with respect to the rejections of all pending claims have been fully considered and are partially persuasive in light of applicant's amendments.

The rejection of claims 6, 33, and 40 under 35 USC 112, second paragraph, has been withdrawn since applicant's amendment corrected the dependencies.

The objections to claims 1, 25, 35, and 46 stand withdrawn in view of applicant's amendments to correct minor formatting and spelling errors.

The rejection of claims 1, 25, and 35 under 35 USC 103(a) in view of Pighin in view of Rowe stand withdrawn in view of applicant's amendments.

The rejection of claims 1-2, 6-7, 10, 13-29, 33-39, and 43-47 under 35 USC 103(a) over Pighin in view of Simon does not stand withdrawn.

As a preliminary matter, examiner wants to point out that the recitation "rendering a single frame of a synthesized image" is an intended use. As such, since the recitation occurs in the preamble, it need not be given patentable weight. Additionally, the CAFC has held repeatedly that article adjectives mean "one or more" (most recently *Scanner Technologies Corp. v. ICOS Vision Systems Corp.*, 70 USPQ2d 1900. "The indefinite article "a" or "an" carries meaning of "one or more" in open-ended claims containing transitional phrases "comprising," and unless claim is specific to number of elements, article 'a' receives singular interpretation only in rare circumstances in which patentee evinces clear intent to so limit article.") In the instant case, examiner points out that

"synthesized image" is construed and defined to be an image having multiple frames. Multiple single frames are therefore a 'synthesized image'. Finally, examiner maintains that the preamble does not have to be given patentable weight in these circumstances, and that the present case does not rise to the level of "rare circumstances" required by the federal circuit, since the end result is a plurality of single frames and applicant's specification is directed towards animation.

As a second point, examiner points out that applicant's arguments are primarily directed to "blending along boundaries of adjacent subregions that do not have discontinuities in texture," where 'adjacent subregions' share a common boundary. Applicant contends, "A broad meaning is implied from blending from blending discussed in Pighin et al and Simon et al…" (Page 4, Remarks, applicant's number 14). However, applicant is arguing for a narrow definition of blending. For example, if two regions having the same skin tone were to be blended, it is unlikely that there would be significant discontinuities. Finally, the specification is not precisely clear what level of distortion (or lack thereof) constitutes "without discontinuities," therefore examiner is construing it broadly (since any practical system will include at least some (negligible) distortion from lack of accuracy and/or precision).

Pighin clearly is directed towards generating blending without discontinuities; Pighin teaches: "We employ our system not only for creating realistic face models, but also for performing realistic transitions between different expressions. One advantage of our technique, compared to more traditional animatable models with a single texture maps, is that we can capture the subtle changes in illumination and appearance (e.g.

facial creases) that occur as the face is deformed. This degree of realism is difficult to achieve, even with physically based models...We develop a morphing techniques that allows for different regions of the face to have different "percentages" or "mixing proportions" of facial expressions. We also introduce a painting interface, which allows users to locally add in a little bit of an expression to an existing composition expression. We believe that these novel methods for expression generation and animation maybe more natural for the average user than more traditional animation systems..." (Section 1, page 2) Clearly, Pighin contemplates this particular problem and attempts to provide more realistic transitions between faces, which would constitute generating frames of images. Additionally, since the painting interface is present, one can fairly draw the inference that such adjustments are done on a frame-by-frame basis. Therefore, the concept of minimizing distortion or discontinuities is clearly contemplated (e.g. either a view-independent texture map or plural view-dependent texture maps are extracted (section 1, page 2).

See page 3, section 3 and 3.1, as additional and more specific proof, wherein weight maps are generated for combining the different textures.

Specifically, see item (2) in the list of important considerations when defining a weight map: "***Smoothness*: the weight map should vary smoothly, in order to ensure a SEAMLESS blend between different input images.**" Clearly, a "seamless" blend would constitute the recited "without discontinuities" blend.

Pighin further teaches blending regions, such as blending eye, teeth, ears, and hair separately, while taking into account the shadowing by the eyelids and lips, etc (3.4).

More importantly, note the Blend Specification section (4.2). Blending weights can be set based on a regional blend. Specifically, prior results indicate that partitioning the face into three regions (forehead, eyes, and lower part of the face) and then further subdividing it vertically on a line down the center into a total of six regions results in a set of coherent regions that are linked. Pighin *et al* clearly partitions the face into several (softly feathered, e.g. blended, wherein that technique (feathering) is known in the art to produce images that appear seamless).

In the case of section 4.2 (and the discussion of the user-modifiable paint technique), the idea is to create seamless integration between the different regions. Clearly, the regions as described in section 4.2 constitute 'adjacent regions having common boundaries' and there is the suggestion in 3.4 to render the eyes, teeth, ears, and hair as separate regions, which would be adjacent and have common boundaries with the recited regions in section 4.2 as per the recited claim.

Finally, it would have been obvious that the goals of creating the weight maps in section 3.1 (e.g. **smoothness** and seamless blending) would apply to blending the various regions, wherein feathering is one example of a technique that can be employed to blend boundaries and regions. Inherently, feathering regions that are adjacent and share a common boundary is a form of blending, and clearly the goal is to produce boundaries free of discontinuities, in other words **seamless blending**.

Examiner next refers to Simon [0049], wherein faces are segmented into regions. However, there is a clear teaching of spatial feathering, alpha masks, and [texture] enhancement filters to ensure smooth transitions between regions that have been spatially enhanced and those that have not.

Simon therefore teaches the division of the face into adjacent regions. As noted above, both regions clearly express the point that the regions are adjacent to each other.

Again, there is the consistent emphasis across both Pighin and Simon of the desire to have **smooth transitions** (Simon) and **seamless blending** (Pighin) by utilizing feathering, alpha masks, enhancement filters, and the like. Examiner submits again that both references clearly speak to the desire to have regions that are blended in a manner that appears natural, e.g. "without discontinuities." The definition of "discontinuity" is "a lack of continuity; irregularity; a break or a gap." Therefore, this concept is synonymous with and comparable to the desired and stated end result of Pighin and Simon – an image that does not have any visible discontinuities between regions.

As a third point, Examiner contends that "subregions defined adjacent to each other" means that such subregions share common boundaries, and applicant is reciting an inherent property.

To return to a brief discussion of how applicant defines blending, it is instructive to point out that applicant's specification describes it as an outgrowth of Pighin's work (see 12:24-13:20 for example). Next, consider that applicant recites that the goal of the present invention is to produce a system that models expression wrinkles and the like to produce photorealistic image expression (see 12:5-13). Pighin has the same goals as described on page 2, section 1. "One advantage of our technique ... is that we can capture the subtle changes in illumination and appearance (e.g. facial creases) that occur as the face is deformed..."

The techniques described in applicant's specification are comparable to those described by Pighin (for example, applicant states (15:8-20) that the face is broken into regions (Pighin section 4.2, regional blend, for example) including subregions for the teeth (Pighin 3.4)).

Specifically, applicant defines blending (19:18→): "Blending can take many forms. In one embodiment, a fade-in fade-out blending technique is used along the subregion boundaries. In one implementation, a weight map is used to facilitate the blending..."

Compare this to Pighin – feathering is used in regional blending. Feathering is a technique wherein the transparency (e.g. alpha) values of a region are lowered towards the edge so that a seamless blend is achieved. In the case of two images being merged, the transparency of one image is decreased towards the boundary until alpha goes to zero (e.g. feathering). Pighin clearly uses feathering (section 4.2) and weight maps (section 3.1, section 3.3, particularly section 4.1) and the use is suggested for

regional blending as well, wherein it is desirous for all pixels in the same region to have the same weight, wherein clearly this involves the use of weight maps.

Therefore, based on applicant's own definitions, the blending techniques described in the specification are broad and are precisely those described in the Pighin reference.

Claim Rejections - 35 USC § 101

35 U.S.C. 101 reads as follows:

> Whoever invents or discovers any new and useful process, machine, manufacture, or composition of matter, or any new and useful improvement thereof, may obtain a patent therefor, subject to the conditions and requirements of this title.

Claims 1-2, 25, 35, 6-7, 10, 13-29, 33-39, and 43-47 are rejected under 35 U.S.C. 101 because they fail to recite a concrete, practical, and tangible end result. The claims recite 'generating the selecting image for the frame...' but there is no positive recitation of a practical, tangible outcome such as displaying the generated image. See Interim Guidelines, Annex B(ii) among other locations.

Claim Rejections - 35 USC § 112

The following is a quotation of the second paragraph of 35 U.S.C. 112:

> The specification shall conclude with one or more claims particularly pointing out and distinctly claiming the subject matter which the applicant regards as his invention.

Claims 1, 25, and 35 are rejected under 35 U.S.C. 112, second paragraph, as being indefinite for failing to particularly point out and distinctly claim the subject matter which applicant regards as the invention.

Specifically, the claims recite the limitation "blending ... without discontinuities." See the above discussion in the Response to Arguments. Any practical system will

have some amount of discontinuities because of hardware limitations on precision and accuracy. The specification does not provide any standard for measuring "without discontinuities" and one of ordinary skill in the art would not know the standard required to meet the definition in the claims.

Claims 2, 6-7, 10, 13-29, 33-39, and 43-47 are rejected as not correcting the deficiencies of their parent claim(s).

Claim Rejections - 35 USC § 103

The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

> (a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negatived by the manner in which the invention was made.

The factual inquiries set forth in *Graham* **v.** *John Deere Co.*, 383 U.S. 1, 148 USPQ 459 (1966), that are applied for establishing a background for determining obviousness under 35 U.S.C. 103(a) are summarized as follows:

1. Determining the scope and contents of the prior art.
2. Ascertaining the differences between the prior art and the claims at issue.
3. Resolving the level of ordinary skill in the pertinent art.
4. Considering objective evidence present in the application indicating obviousness or nonobviousness.

Claims 1-2, 7, 24-25, and 35 are rejected under 35 U.S.C. 103(a) as being unpatentable over Pighin et al ("Synthesizing Realistic Facial Expressions from Photographs,") in view of Simon et al (US PGPub 2003/0223622 A1) and Lanitis et al (A. Lanitis, C. J. Taylor, and T. F. Cootes, "Automatic interpretation and coding of face

images using flexible models," incorporated by reference into the Simon reference, [0049]).

As to claims 1, 25, and 35,

A computer implemented method for rendering a single frame of a synthesized image, comprising: (Pighin teaches a computer-implemented method for synthesizing and rendering images – see Figure 4)

-Generating a geometric component corresponding to a selected image for the frame based on identified feature points from a set of representative images (Pighin section 2, page 2, "We…recover the 3D coordinates of a set of feature points on the face…" where this is of a set of representative images, note Figure 4, where 2 exemplary expressions of the actor were captured – Pighin 4 shows generating a geometric component, where clearly teeth and eyes must be generated (section 3.4, page 5) separately, the base expressions are used to synthesize a final one), **where each image of the set has the identified feature points, and wherein the geometric component is a dimensional vector of feature point positions;** and (Pighin clearly suggests the incorporation of automatic modeling, where the system would find features automatically). Finally Pighin clearly divides facial images into the face, eyes, teeth, and ears (section 3.4, page 5), so the idea of dividing facial images into regions in a set of representative images is clearly taught. Also, Pighin clearly shows (for example, the database of actor / individual expressions capture and shown in Figure 4) that clearly a set of representative exists, and that each image has the same features – they simply move. Clearly the coordinates of those points in 3D would constitute a 'dimensional

vector')(Simon clearly teaches [0011] that images are segmented into different regions [0049], where a plurality of face images can be used, and that these different regions include skin, eyes, eyebrows, nose, mouth, neck, and hair regions [0011,0044]– see Figure 4 as an example [0050]. Therefore, generating one (1) resultant image would constitute 'generating a single frame of a synthesized image.' Clearly, Simon generates new resultant regions based on the results of the image retouching process [0058], where the filters are applied to each sub-image and then the final results are displayed. The system of Simon also generates new features or changed features based on changes in textures [0060]. More importantly, Simon clearly teaches that such regions 'feature maps' are generated and/or refined [0080-0082] for the resultant portions of the face. Therefore, each image of the set will have the recited feature points)

-Generating the selected image for the frame (Pighin Figure 4 as explained above) **from a composite of the set of representative images** (More specifically, Pighin breaks the face up into multiple regions (see section 3.4 as a basis, e.g. eyes, ears, mouth, and hair are separate regions and are processed separately). Then in section 4.2, it is clear that the face is divided up into coherent regions analogous to the six-region model discussed and then the blends are performed for each region. This clearly constitutes 'a composite of the set of representative images'. Additionally, Pighin Figure 4, contains images – "surprised," left and "sad," center – which are 'a composite of the set of representative images, since it is generated by a global blend – see caption on Figure 4. Pighin suggests a multi-way blend in section 4.1 on page 6, over the original

set of representative images. In any case, blending specific regions would show that there exists a set of composite of set of representative images for that specific region or element, specifically since certain things are rendered separately for the view-dependent mode (3.4), etc.)(Simon blends the new image region portion with the original image region portions – see Figure 13) **based on the geometric component** (Pighin teaches facial features can be blended based on regional blends in section 4.2, where the mixing proportions for each region varies – see Figure 5 caption and explanation in section 4.2, where clearly a region or sub-portion (e.g. eyes, forehead, nose, and/or the like) would be contemplated)(Simon produces the output image on a per-feature or per-region basis), **wherein the selected image and each of the set of representative images comprises a plurality of subregions defined adjacent to each other wherein adjacent subregions share a common boundary** (Pighin shows that subregions (which correspond to the geometric components) do exist next to each other, since clearly the eyes exist next to the skin portion of the face)(Simon clearly shows in Figure 4 that features such as eyes, nose, eyebrows, hair, etc, exists adjacent to each other – eyebrows are adjacent to eyes, for example), **and wherein generating a geometric component is performed for each subregion;** (Simon obviously tracks each subregion and generates a new version of it if requested – as explained above [0080-0082], Figure 13, and the like) **and wherein the composite of the set of representative images is based on the corresponding geometric component for each subregion**, (Pighin region blend in section 4.2 and Figure 5, which would (with a multi-way blend, as in section 4.1) generate a composite of the set of representative

images)(Simon Figure 13, blending of region of original image and altered or enhanced region of original image, where this clearly represents) **and the selected image includes a synthesized subregion for each subregion based on the composite by blending at least some boundaries between adjacent subregions of the selected image without discontinuities in texture in order to generate the selected image.** (Pighin clearly shows how such images are generated adjacent to each other as explained above, where the blending referred to is done with respect to textures – see the suggestion in section 7 (Future work, page 8): "To improve the quality of the composite textures, we could locally warp each **component** texture (and weight) map before blending". Clearly the idea of blending between adjacent subregions is contemplated or suggested. Pighin synthesizes resultant images as in Figures 4 and 7-8. More specifically, Pighin talks about appropriate criteria for blending (e.g. the desired end results in the 4 –part list described in section 3.1 and emphasizes seamless blending / smoothness as a desirable characteristic. In section 4.2, regional blending is discussed and feathering techniques are described to blend regions of the face that are adjacent to each other (wherein Pighin has divided the face into coherent regions next to each other). Given the regional divisions described in section 4.2, it is further noted that the concept of individual regions for eyes, ears, mouth, and hair are contemplated in section 3.4, wherein the regions for synthesizing these regions are still logical for their separate synthesis as explained in section 4.2)(Simon very clearly feathers the regional definition masks to create alpha masks [0049]. These feathered binary masks and alpha masks are used in blending operation: "Feathering binary masks and applying the

resulting alpha masks in blending operation ensure smooth transitions between regions that have and have not been enhanced. To generate alpha masks the binary masks are feathered by blurring the binary masks with a blurring function where the blur radius is chosen based upon the size of the face ..." Therefore this would clearly constitute 'synthesizing subregions' that are adjacent to each other – note previous discussion in first clause, where clearly blending is done at the boundaries via the alpha masks to avoid discontinuities in texture. Indeed, the point of using alpha masks is such that there will be a continuous texture and there will not be abrupt artifacts that can occur (see Pighin, Future Work, section 7, where it is stated that to improve issues regarding textures ghosting and blurring, local texture warps and blends would improve the situation (in addition to the feathering described in section 4.2 and the desired smoothness / seamless blending described in section 3.1), which Simon does perform, as cited above).)

In summary, Pighin does not expressly teach that the feature points are dimensional vectors, but the Simon reference supplies this teaching.

Pighin shows blending with a regional implementation, wherein Simon explains why this can be more beneficial when the desire is to perform enhancement processing on only certain images on a face or the like.

It would have been obvious to one of ordinary skill in the art at the time the invention was made to combine Pighin and Simon for at least the fact that Simon

provides automatic registration and segmentation of images into regions [0011-0012, 0040, and the like], where Pighin suggests adding this feature under 'Automatic modeling' and the fact that Simon provides additional methods of blending local regions together effectively, where Pighin described some methods on how local blending could take place and how to effectively avoid discontinuities (as noted in the "Improved registration" section under Future Work). It is therefore clear what references teach which limitations.

Specifically, see the Remarks to Arguments section above for a discussion of the specific blending issues, and the specifically cited sections of the references in the rejection above.

As to claim 2, Pighin clearly **calculates a plurality of values different from the feature points** by calculating texture maps and weight maps, as explained in sections 3 – 3.2, **wherein a value of the plurality of values is associated with each representative image** since each of the model face images in Pighin has its own texture maps and weight maps, and the **plurality of values are used to composite the set of representative images** where Pighin uses the underlying images to perform global, multi-way, or localized blends – see Figures 4-7 and pages 6-7, and Simon teaches compositing a set of representative images (e.g. two, the original and the "enhanced" version).

As to claim 7, the feature points in Simon correspond to two-dimensional images, since these can be taken with a single camera and do not involve complex efforts –

further Pighin teaches – as in Figure 5 and page 7 – that the user expressions take place as part of a set of photographs.

As to claim 24, examiner submits that Pighin uses a representative face model, where each of the set of representative image is aligned to it (all of page 2, particularly section 2), where that would constitute an underlying reference image (which would be three-dimensional).

Claims 6, 10, 13-14, 26, and 36 are rejected under 35 USC 103(a) as unpatentable over Pighin in view of Simon as applied to claim 1 above, and further in view of Cosatto et al ("Photorealistic Talking Heads from Image Samples", cited on previous 892)

As to claim 6, examiner submits that Pighin implicitly suggests **that one synthesized subregion is based on a quantity of a set of representatives different than another synthesized region** where Pighin discusses localized blending for expression synthesis in section 4 on page 6.

However, Pighin and Simon do not expressly teach the above. It is submitted that Pighin teaches generating animated transition frames Figure 6 and section 5 on page 7.

Cosatto teaches this limitation (see second paragraph below) and is an analogous art, as explained in this paragraph. In section 1, page 152, that in the first step, image samples of facial parts are generated and results in a database of facial

parts. Pages 153-154, section III, teach the methods of how this is done, and how the hierarchy of parts and samples are obtained and subsequently ordered. Section IV on page 154 states that the first step in the process is measuring the face to determine the location of certain facial points (e.g. the recited feature points), which correspond to the "identified feature points" above. The "set of representative images" is the video recorded in section I; all faces would have the same general set of features, e.g. eyes, nose, et cetera. The system of Cosatto clearly synthesizes a geometric component, e.g. synthetic video, with specific emphasis on for example the mouth, section V-B (pages 159-161) with other facial parts discussed in section V-D (page 161), which clearly constitutes "generating a geometric component", and the selected image is simply one frame of video wherein the synthesized face is saying something (e.g. see section V-B).

Since the system of Cosatto is intended primarily for synthesizing the mouth region to create natural-appearing speech, there would obviously be more samples of the mouth region than of other regions, particularly in the database of parts, as that is derived from all the video-recorded phonemes. Therefore, either Cosatto implicitly teaches it or it is a trivially obvious variant, and it would be obvious to modify for the reasons set forth immediately above, and on page 59, section H, it is stated the mouth database is larger than those of other features and an absolute size provided.

It would have been obvious to one of ordinary skill in the art at the time the invention was made to combine the system of Cosatto with Pighin/Simon such that their

system could incorporate pre-recorded speech or video (see for example the suggested Future Work section on pages 8-9 in Pighin) and generate photo-realistic results.

As to claim 10, this is an obvious variant of claim 7, wherein Cosatto in section 2 (pages 152-154, emphasis on page 153) states that three-dimensional images using 3-D scanners are common in the art and in prior work. As further discussed in section IV-A on page 154, feature points on the face are measured in 3-D. Therefore, it would be obvious that the feature points could be on a three-dimensional image and it would be obvious to modify the system of Cosatto to use three-dimensional images for the reasons set forth above. Motivation and rationale is taken from the rejection to claim 6 above.

As to claim 13, Cosatto teaches in section A on page 155 that "Knowing the position of a few points in the face allows to recover the 3-D head pose from 2-D images", where this clearly justifies that examiner's contention that the a few key feature points are used to extract the position of other feature points, see for example section VI-D on page 157. Section V-B on pages 159-160 clearly teaches how knowledge of a few points allows synthesis of a great many essential feature points on the mouth, which is the key feature. Motivation and rationale is taken from the rejection to claim 6 above.

As to claim 14, Cosatto teaches that obviously feature points are grouped in sets by different regions of the face – see page 154, sections III-1 through III-4 and Fig. 1 or of the synthesized image – see page 161, sections D and E. Finding the position of one feature point on for example the mouth (see section V-B and V-C, particularly page 160)

allows the calculation of the shift in other portions of the images, e.g. where the changes in position between one frame and another of the synthesized image are minimized to get more natural appearance (e.g. Figure 7 on page 160), which prima facie tracks change in position of feature points within the mouth region so as to be able to calculate the path that involves the least change in position for Viterbi optimization, and the details on feature point location and tracking are found in sections III-D and III-E, particularly section III-D. Thusly, Cosatto teaches all the limitations. Motivation and rationale is taken from the rejection to claim 6 above.

As to claim 26, this claim is essentially a duplicate of claim 14, with the difference that Cosatto teaches that the feature points are grouped in sets according to the region of the face, e.g. the hierarchical database shown on page 154, and the rest of the limitations are taught in the rejection to claim 14, which is herein incorporated by reference in its entirety. Motivation and combination are taken from claim 6 above.

As to claim 36, this claim is a substantial duplicate of claim 26, with that rejection herein incorporated by reference; motivation and combination is from claim 26 above.

Claims 15-23, 27-29, 33-34, 37-39, and 43 are rejected under 35 U.S.C. 103(a) as being unpatentable over Pighin, Simon, and Cosatto as applied to claim 14 above, and further in view of Chai et al (Chai et al. "Vision-based control of 3D animation".)

As to claim 15, Pighin, Simon, and Cosatto do not expressly teach the limitation of using PCA to track position. Chai teaches the use of PCA on pages 200-201 for

example with emphasis on sections 4.2 and 4.3, where it is taught that using PCA, the motion frames are broken down into linear subspaces and motion is tracked in that way. On page 200, section 4.3 it clearly discloses that a database of motion is kept, which would be similar to the database of images in Pighin, Simon, and Cosatto. The database of motion would be with respect to each linear subspace, which obviously could be the different facial regions of Cosatto – that is, the positional changes in motion of the images in the database of Cosatto for facial regions could be found using the PCA techniques of Chai. Therefore, It would have been obvious to one having ordinary skill in the art at the time the invention was made to combine the PCA of Chai with the motion tracking and splitting of the face into different regions of Cosatto and Pighin/Simon for the reasons set forth above, as using PCA allows faster computation times for motion detection and improves temporal coherency (pg. 200- section 4.4 for example).

As to claim 16, it would have been obvious that given that Cosatto tracked motion using overall feature points on the face (e.g. section III-A page 155 or section D page 161) and that Cosatto also tracked feature points within the mouth subset in order to assure more natural appearing features as the difference between each pose was minimized via Viterbi optimization on page 160, Figure 7. Obviously, overall changes in head position would tracked via the main feature points and determining the necessarily positional changes in the mouth (besides those necessitation by normal motion of talking) would be based on the positional changes in the larger set of feature points on the face itself, e.g. any necessary translational or rotational movement of the overall

head for example. Since only the parent references are utilized, no separate motivation or combination is required and that from the rejection to the parent claim is herein incorporated by reference.

As to claim 17, the system of Cosatto has a hierarchical database structure of feature parts, see for example section III, page 154, items 1-3, particularly Item I, titled "Hierarchy of Parts." Since only the Cosatto reference is utilized, no separate motivation or combination is required and that from the rejection to the parent claim is herein incorporated by reference.

As to claim 18, the system of Pighin/Simon and Cosatto does not expressly teach this limitation, insofar as it teaches tracking feature points of the user when the data for the initial sets is recorded, but it does not expressly perform the recited details, although it does monitor feature points of a user. The system of Chai performs the recited limitations, in that it consists of a video camera that monitors the face of a user and generates an image of an avatar making similar facial movements, see for example Fig. 1 on page 193, the caption specifies that users act out the motion in front of a single-view camera, and that the avatars have controlled facial movements similar to those of the user with texture mapped models (see section 1, left side of page 194, and Figure 2 on page 195, and the captions on it). The system of Chai further tracks feature points of the user (section 2.1, page 196) on the face and moves the avatar as the user moves (see section 1.2 on page 196, where motion data and head motion are separated from facial deformations and then both are applied to the avatar in separate passes). Obviously, the generated avatars of Chai (Figs. 1 and 2 for example) have separate

components of the face, or it would be obvious to use the separate components of Cosatto (+Pighin/Simon) for the face, and to utilize the motion tracking and facial deformation techniques of Chai described above. It would have been obvious to one having ordinary skill in the art at the time the invention was made to combine the systems of Cosatto (and Pighin/Simon) and Chai, since Chai would allow any user to control the facial expressions of an avatar in addition to overlaying audio text and simulating real speech – the facial techniques would allow better synchronization of voice and facial movements in for example the avatars, and would allow even an unskilled user to adequately control facial motions (see section 1, pages 193-194).

As to claim 19, Pighin, Simon, and Cosatto teaches in (Cosatto) section A on page 155 that "Knowing the position of a few points in the face allows to recover the 3-D head pose from 2-D images", where this clearly justifies that examiner's contention that the a few key feature points are used to extract the position of other feature points, see for example section VI-D on page 157. Section V-B on pages 159-160 clearly teaches how knowledge of a few points allows synthesis of a great many essential feature points on the mouth, which is the key feature. Since only the primary references are utilized, no separate motivation or combination is required and that from the rejection to the parent claim is herein incorporated by reference.

As to claim 20, references Pighin, Simon, and Cosatto do not expressly teach this limitation, insofar as it does teach rendering an image of a speaking human being with the identified feature points on it (see for example Fig. 8, and facial locations are tracked by feature points as illustrated by Fig. 4, where the control points are noted.

However, Chai teaches on page 196 in the "initialization" section that the user can select the control points, for which it would be an obvious modification to allow the user to control the movement of a feature point. Also, since the system of Chai (for example, see caption on Fig. 1 on the first page) teaches that the avatar moves in response to user facial and head movements, this also constitutes "receiving information indicative of a user moving a feature point". It would have been obvious to one having ordinary skill in the art at the time the invention was made to combine the systems of Pighin/Simon/Cosatto and Chai, since Chai would allow any user to control the facial expressions of an avatar in addition to overlaying audio text and simulating real speech – the facial techniques would allow better synchronization of voice and facial movements in for example the avatars, and would allow even an unskilled user to adequately control facial motions (see section 1, pages 193-194).

As to claim 21, this claim is a substantial duplicate of claim 16; the rejection to that claim is herein incorporated by reference in its entirety, along with motivation and combination.

As to claims 22 and 23, Pighin, Simon, and Cosatto do not expressly teach this limitation, whilst Chai teaches in Fig. 1 on page 193 that the user can control or select the facial expression by making the desired expression on their own face, e.g. two separate facial expressions are shown in the leftmost column, and in the rightmost column the avatars are shown depicting those facial expressions. Motivation and combination is incorporated by reference from claim 20 above.

As to claim 27, this claim is a substantial duplicate of claim 15, with the only difference that the database of representative images of Cosatto is substituted for the motion database of Chai. Chai teaches a motion database on page 194 on the left side of the page in section 1 and in the caption to Figure 2 on page 195, Chai teaches that the motion database can be used to synthesize expressions. The rest of the limitations are taught in the rejection to claim 15, which are herein incorporated by reference in its entirety; motivation and combination are taken from claim 6 above.

As to claim 28, this claim is a substantial duplicate of claim 16, with that rejection herein incorporated by reference; motivation and combination is from the rejection of claim 27 above.

As to claim 29, this claim is a substantial duplicate of claim 17, with that rejection herein incorporated by reference; motivation and combination is from the rejection of claim 27 above.

As to claim 33, this claim is a substantial duplicate of claim 6, the rejection to which is incorporated herein by reference. Since only the primary reference is utilized, no separate motivation or combination is required and that from the rejection to the parent claim is herein incorporated by reference.

As to claim 34, Pighin/Simon/Cosatto do not expressly teach this limitation, whilst the system of Chai performs the recited limitations, in that it consists of a video camera that monitors the face of a user and generates an image of an avatar making similar facial movements, see for example Fig. 1 on page 193, the caption specifies that users act out the motion in front of a single-view camera, and that the avatars have controlled

facial movements similar to those of the user with texture mapped models (see section 1, left side of page 194, and Figure 2 on page 195, and the captions on it). Motivation and combination are taken from the parent claim, e.g. 25 and herein incorporated by reference.

As to claim 37, this claim is a substantial duplicate of claim 27, with that rejection herein incorporated by reference; motivation and combination is from claim 27 above.

As to claim 38, this claim is a substantial duplicate of claim 28, with that rejection herein incorporated by reference; motivation and combination is from claim 27 above.

As to claim 39, this claim is a substantial duplicate of claim 29, with that rejection herein incorporated by reference; motivation and combination is from claim 27 above.

As to claim 43, this claim is a substantial duplicate of claim 33, with that rejection herein incorporated by reference; motivation and combination is from claim 27 above.

Claims 44 and 47 are rejected under 35 USC 103(a) as unpatentable over Pighin and Simon as applied to claim 1 above, and further in view of Nielsen (US 6,591,011 B1).

As to claim 44, Pighin and Simon do not expressly teach this limitation. Nielsen teaches a method of synthesizing images from a plurality of base images (Abstract), where the system is capable of adjusting tiles or images that have been rotated, transposed, or mirrored to a common frame of reference (Abstract, Figures 14-17B), where image searching and remapping can be done in linear program form using convex hulls (see 18:35-66), where such allow a logarithmic computation cost, which is

clearly lower than that of Pighin. It therefore would have been obvious to one of ordinary skill in the art at the time the invention was made to modify Pighin to utilize the convex hull matching method to speed the matching of the images to the underlying model as in page 2, section 2, and the like.

As to claim 47, as noted above in the rejection to claim 44, which is incorporated by reference (and a convex hull is clearly a convex combination of geometry), where clearly Simon and Pighin generate image coefficients, and so does Nielsen. These coefficients are related to the set of representative images and could be typified by e.g. the weight maps of Pighin as discussed above and in section 3. Motivation and combination is taken from the rejection to claim 44 above.

Claim 45 is rejected under 35 USC 103(a) as unpatentable over Pighin and Simon as applied to claim 1 above, and further in view of Stewart et al (US PGPub 2003/0190091 A1).

As to claim 45, Pighin and Simon do not expressly teach this limitation. Stewart teaches the use of an objective function that uses constraints to perform faster image registration to an underlying model and the like [0094], where feature points are used to do so [0044,0122, and the like]. The process and use of objective functions is summarized in [0011-0015]. It would have been obvious to one of ordinary skill in the art at the time the invention was made to modify the system of Pighin to utilize the improved registration methodology of Stewart, since it is faster (for example, [0023-0024]).

Claim 46 is rejected under 35 USC 103(a) as unpatentable over Pighin in view of Simon and Stewart as applied to claim 45 above, and further in view of Fogel et al (US 5,991,459).

As to claim 46, Pighin, Simon, and Stewart do not expressly teach this limitation. Stewart clearly teaches the use of linear programming and linear constraints as explained above, but does not teach that the objective function is a positive semi definite quadratic form with linear constraints. Fogel teaches the use of such functions and such constraints in 25:14-26:50 in the context of image registration between various frames. Therefore, it would have been obvious to one of ordinary skill in the art at the time the invention was made to modify the system of Stewart as above to utilize the semi-definite quadratic forms and linear constraints of Fogel because it allows the addition of further constraints that shrink the solution space and decrease computational time (24:30-28:50).

Conclusion

Any inquiry concerning this communication or earlier communications from the examiner should be directed to Eric Woods whose telephone number is 571-272-7775. The examiner can normally be reached on M-F 7:30-5:00.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Ulka Chauhan can be reached on 571-272-7782. The fax phone number for the organization where this application or proceeding is assigned is 571-273-8300.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see http://pair-direct.uspto.gov. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free). If you would like assistance from a USPTO Customer Service Representative or access to the automated information system, call 800-786-9199 (IN USA OR CANADA) or 571-272-1000.

Eric Woods

February 12, 2007

U Chauhan
ULKA CHAUHAN
SUPERVISORY PATENT EXAMINER